

THE APPLICATION OF ANALYSIS OF VARIANCE TO DIFFERENT IRON-BASED SUPERCONDUCTOR CRITICAL TEMPERATURE MODELING

Atirat Maksuwan*, Arpapong Changjan and Phanuchai Pramuanl

Received: January 15, 2023; Revised: May 19, 2023; Accepted: June 05, 2023

Abstract

Recently, Iron-based superconductors have shown promising properties of high critical temperature and high upper critical fields, which are prerequisites for applications in highfield magnets. Extensive research has been conducted on a modeling approach that contributes to predicting doped Iron-based superconductor critical temperature from structural and topological parameters. Statistical significance of differences in modeling approaches requires studies that can reliably distinguish between systematic approach effects and errors resulting from modeling approach variation. In this work, we introduce analysis of variance (ANOVA) to assess the statistical significance of differences in modeling approach variation. Comparisons of obtained results with Iron-based modeling approach variation data of support vector machine (SVM) and linear regression with natural logarithm transformation (LRNLT) were presented.

Keywords: Analysis of variance (ANOVA); Critical temperature; Iron-based superconductor; Statistical significance

Introduction

The trend of accidental discovery of superconductivity in mercury in 1911 in Leiden University by the Dutch Physicist, K. Onnes extended to the recently discovered iron based superconductors of high critical superconducting temperature T_c (K). Heike Kamerlingh Onnes was investigating the behavior of the residual resistivity of high purity mercury when accidentally discovered superconductivity. More recently; Hideo Honoso of Toyko Institute of technology desired to fabricate transparent semiconductor, he stumbled on

a new type of superconducting material; the Iron based (Kamihara *et al.*, 2008).

In iron-based superconductors, high T_c (K) superconductivity can be induced and tuned by various doping methods and dopant levels. In general, superconductivity is induced when antiferromagnetism is diminished by carrier doping, structural modification under external pressure, or chemical pressure via isovalent substitution. Therefore, it is important to investigate the relationship between doping mechanisms and

Department of Environmental Technology and Applied Science, Faculty of Science and Technology, Pathumwan Institute of Technology, Bangkok 10330, Thailand. Tel. 0-2104-9099; E-mail: atirat.maksuwan@gmail.com; otto_sinkronity@yahoo.com; Phanuchai.p@gmail.com

* Corresponding author

DOI: <https://doi.org/10.55766/sujst-2023-04-e02609>

Suranaree J. Sci. Technol. 30(4):030128(1-9)

changes in critical temperature T_C (K). T_C (K) is influenced by several factors, including topological lattice parameters disorders and electronic structures. Generally, high- T_C superconductivity can be induced and tuned by varying dopant types and levels, where antiferromagnetism is diminished by carrier doping, structural modifications under external pressure, or chemical pressure via isovalent substitutions.

Combined with different doping mechanisms, a large variety of Fe-based superconductors were synthesized, such as AFeSe (A = alkali), AeFe₂As₂ (Ae = alkali-earth), and LnOFeAs (Ln = lanthanide). Previous studies have indicated that changes in structural and topological lattice parameters constants reflect the expansion or contraction of the interlayer spacing of FeAs layers and local geometry of the FePn(Ch)₄ (Pn = pnictide, Ch = chalcogenides) tetrahedron (Chen *et al.*, 2014). So, it is of great importance to investigate the correlation between topological lattice parameters and T_C (K), which are much easier to determine experimentally as compared to other physical parameters (Cai *et al.*, 2010; 2013; Owolabi *et al.*, 2014; 2015; Hamidieh, 2018).

Recently, data-driven analysis has been proved to facilitate the understanding of the relationship between material structures and performance by finding statistical correlations between physical attributes and property parameters “Reference of research articles from 15th-43rd” shows sources of experimental data that is obtained from Zhang and Xu (2021).

Therefore, the success of computational techniques directly relate to topological lattice parameters and the T_C (K) paves a significant way for quick and accurate estimation of T_C (K) and eventually eases the usual highly demanding experimental procedures that involve the use of expensive cryostat (Owolabi *et al.*, 2016).

In this work, we introduce analysis of variance (ANOVA) to assess the statistical significance of differences in modeling approach variation. To demonstrate the outstanding efficiency of the ANOVA results. The comparisons of obtained results with iron-based modeling approach variation data of support vector machine (SVM) and linear regression with natural logarithm transformation (LRNLT) were presented.

Materials and Methods

Brief Description of Analysis of Variance (ANOVA)

ANOVA is used to determine significance using the ratio of variance estimates from sample

means and sample values (Krzywinski and Altman, 2014). Their ratio or test statistic F can be used between group mean squares (MS_B) and within group mean squares (MS_W), which are calculated from between treatment sum of squares (SS_B) and within treatment sum of squares (SS_W) weighted by degrees of freedom (d.f.). The degrees of freedom (d.f.) shows the number of independent deviations in the treatment sum of squares (SS); the deviations are not all independent because deviations of a quantity from its own mean must sum to zero. The ANOVA null hypothesis is that all samples are from the same distribution and have equal means. Under this null, between group variation of sample means and within group variation of sample values are predictably related. Their ratio can be used as F and t tests statistic, which will be larger than expected in the presence of treatment effects. We introduced the idea that variance can be partitioned within group variance was interpreted as experimental error and between group variance as structural and topological parameters variation. In one way ANOVA, the relevant quantities are within group variance and between group variance. Within group variance corresponds to variance in the sample after other sources of variation have been accounted for and represents experimental error. If some sources of error are not accounted for (e.g., structural and topological parameters variation), within group variance will be inflated. Between group variance is another estimate for within group variance, additionally inflated by average squared deviation of treatment means from the grand mean. As noise within samples drops, a larger fraction variation is allocated to between group variance, and the power of the test improves (Armstrong *et al.*, 2022). This suggests that it is beneficial to decrease within group variance. We can do this through a process to identify and isolate likely sources of sample variability.

Iron (Fe) and Superconductivity

Superconductivity generally sets in with the formation of Cooper pairs. In Fe-based superconductors, spin polarization in ferromagnets reduces the number of superconducting electron pairs and can be noted in a significant suppression of T_C (K). In general, ferromagnetism is a result of the collective behavior of the distributed charges and spins in a system and sensitively depends on lattice symmetries. Fe-based superconductors can be viewed as a two-dimensional (2 D), slightly distorted, square lattice of Fe²⁺ ions embedded between chalcogen or pnictogen ion layers, that shows an antiferromagnetic order in its ground state. Superconductivity arises after suppression of the antiferromagnetic long range order either

by chemical substitution leading to charge carrier doping. For the moment, we will focus on the different structures of Fe-based superconductors that can be noted in the significant suppression of T_C (K). An overview of the different crystallographic structures of Fe-based superconductors that are often designated according to their stoichiometry as ‘11’, ‘111’, ‘1111’ and ‘122’ compounds. In general, distinct bonding regimes are encountered in all structures within the lattice parameters of the tetragonal unit cells a (Å) and c (Å) together with (Fe_2X_2) layers. The chemical bonding is of a mixed covalent/metallic nature. So, iron (Fe) and superconductivity can be viewed as a system and sensitively depending on lattice symmetries. It is

a result of the collective behavior of all structures within the lattice parameters of the tetragonal unit cells a (Å) and c (Å) which can be noted in a significant suppression of T_C (K). In empirical study, the actual data (obtained experimentally) employed in this research work was drawn from literatures (Owolabi *et al.*, 2014). Total number of thirty-three data set was used for both training and testing as presented in Table 1.

Proposed Method

The uniqueness of this research work is to show that it adopts the ANOVA together with assessing the statistical significance of differences in the modeling approach will also enhance

Table 1. The experimental T_C (K) data and superconducting transition temperature predictions from “SVM Estimated T_C (K)”, “LRNLT Estimated T_C (K)” and “LRNLT (C.F.) Estimated T_C (K)”

Index	Sample	a (Å)	c (Å)	Experimental T_C (K)	SVM Estimated T_C (K)	LRNLT Estimated T_C (K)	LRNLT (C.F.) Estimated T_C (K)
1	LaFeAsO _{0.89} F _{0.11}	4.028	8.713	28	22.9	21.13	22.80
2	LaFeAsO _{0.86} F _{0.14}	4.025	8.695	20	21	21.12	22.88
3	LaFe _{0.89} Co _{0.11} AsO	4.035	8.713	14.3	18.7	21.08	23.07
4	SmFeAsO _{0.93} F _{0.07}	3.393	8.482	35	34.6	26.75	12.96
5	KFe ₂ As ₂	3.842	13.838	3.6	4.1	13.43	6.91
6	SrFe _{1.8} Co _{0.2} As ₂	3.928	12.303	19	18.8	17.64	27.48
7	SrFe _{1.4} Ru _{0.6} As ₂	3.992	12.064	19.3	19.4	18.21	20.39
8	SrFe _{1.3} Ru _{0.7} As ₂	4.005	12.009	19.3	19	18.33	18.80
9	SrFe _{1.3} Ru _{0.8} As ₂	4.011	11.984	17.6	17.9	18.39	18.08
10	Ba _{0.89} K _{0.11} Fe ₂ As ₂	3.949	13.088	2.5	3	15.52	28.99
11	Ba _{0.82} K _{0.18} Fe ₂ As ₂	3.937	13.155	25.4	24.9	15.33	28.89
12	Ba _{0.7} K _{0.3} Fe ₂ As ₂	3.919	13.263	36.4	36.8	15.03	27.99
13	Ba _{0.6} K _{0.3} Fe ₂ As ₂	3.915	13.294	38.6	38.1	14.95	27.59
14	Ba _{0.56} K _{0.44} Fe ₂ As ₂	3.907	13.335	36.8	36.1	14.83	26.84
15	Ba _{0.38} K _{0.62} Fe ₂ As ₂	3.887	13.506	29.6	29.3	14.36	22.87
16	Ba _{0.29} K _{0.71} Fe ₂ As ₂	2.88	13.569	14.6	16	14.48	14.22
17	Ba _{0.15} K _{0.85} Fe ₂ As ₂	3.852	13.735	8.9	9.3	13.71	12.36
18	Ba _{0.1} K _{0.9} Fe ₂ As ₂	3.848	13.793	8.9	8.4	13.56	9.75
19	LiFeP	3.692	6.031	6	6.3	12.52	9.45
20	FeNi _{0.02} Se	3.774	5.519	5	8.6	8.70	8.48
21	FeCo _{0.02} Se	3.771	5.522	6	9	8.73	8.36
22	Fe _{1.01} Se	3.768	5.521	8.2	9.3	8.75	8.29
23	Fe _{1.01} Se (Q)	3.771	5.52	9	9	8.72	8.38
24	FeMn _{0.04} Se	3.77	5.522	10	9.1	8.74	8.33
25	FeMn _{0.04} Se (Q)	3.775	5.527	10	8.4	8.74	8.43
26	FeCr _{0.02} Se	3.773	5.524	10.5	8.7	8.73	8.40
27	FeCr _{0.02} Se(Q)	3.767	5.519	10.8	9.6	8.74	8.28
28	FeV _{0.02} Se	3.772	5.5242	9.5	8.8	8.74	8.37
29	FeV _{0.02} Se (Q)	3.772	5.5205	11.2	8.8	8.72	8.40
30	FeV _{0.03} Se (Q)	3.776	5.508	9.2	8.5	8.62	8.64
31	FeV _{0.05} Se (Q)	0	0	0	0.4	0.10	0.13
32	FeT _{0.02} Se	3.773	5.525	8	8.6	8.74	8.39
33	FeV _{0.01} Se (Q)	3.77	5.18	11	11.2	6.77	13.25
	Mean	3.699	8.939	15.22	15.23	13.09	15.04
	Maximum	4.035	13.838	38.6	38.1	26.75	28.99
	Minimum	0	0	0	0.4	0.1	0.13
	Standard Deviation	0.697	3.904	10.64	10.32	5.44	8.29
	Variance	0.485	15.240	113.29	106.41	29.54	68.69
	Correlation with Experimental T_C (K)	0.283	0.520	-	-	-	-
	Root Mean Square Difference (RMSD)	-	-	-	10.81	56.16	48.61
	Mean Bias Difference (MBD)	-	-	-	0.05	-14.00	-1.14

Note: “ a (Å)” and “ c (Å)” are lattice parameters of Iron-based superconductors. “Reference” shows sources of all experimental data in Owolabi *et al.* (2014)

researchers that are incapacitated to acquire equipment to carry out low temperature research as well as, serving as a premise through which the full understanding of superconductivity can be grown. To demonstrate the outstanding efficiency of the ANOVA results. In this work, we compared obtained results with iron-based modeling between support vector machine (SVM) and adopting a simple model (linear regression) through which they developed by using natural logarithm transformation (LRNLT) approach. In addition, the outstanding performance of ANOVA leads to the creation of an optimization correction factor (C.F.) for simple models.

Working Principle of SVM

SVM adopts support vector regression (SV regression) and kernel trick to generate a function that relates the predictors to the target through which unknown targets are predicted. It works on statistical learning theory together with the structural risk minimization principle (Cortes and Vapnik, 1995; Vapnik, 1995). and principle of artificial intelligence (that is, leaning from experiences). During the training period, SV regression takes some input parameters (such as hyperparameter λ , regularization factor C and kernel -option) by which the chosen kernel function generates a relation between the predictor and the target. The hyper-parameter guides SVM to adopt hyper-plane that maximizes the margin as well as minimizes a parameter which has direct influence on likely errors between the actual and the predicted values (Goldman *et al.*, 2007). The SV regression functions that are obtainable in the literature are

$$f(x) = \sum_{i=1}^n (\lambda'_i - \lambda_i) K(x_i - x) + d \quad (1)$$

Where $\lambda'_i \geq 0$, $\lambda'_i \leq C$, $K(x_i - x)$ is the polynomial kernel function, and d kernel parameters.

Working Principle of LRNLT

The multiple linear regression models are built on the same simple linear regression (Goldman *et al.*, 1977; Diaconis and Efron, 1983; Efron and Gong, 1983). However, in addition to the concepts discussed so far for simple linear regression, which is remain applicable, a new concepts set has to be introduced. This discussion will focus on situation where there are two predictor and one outcome variable. With all three variables, three-dimensional figure can be used to visualize data. Models with more predictive variables follow the same principle,

but difficult to visualize. We consider multiple regression and linear relation as follows,

$$Y = XB + V \quad (2)$$

in which,

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1m} \\ 1 & x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix},$$

$$B = \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_m \end{bmatrix}, V = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_m \end{bmatrix}$$

We develop a multiple linear regression model to predicting doped iron-based superconductor critical temperature from structural and topological parameters by using linear regression with natural logarithmic transformation (LRNLT). Start by considering the nonlinear regression problem is as follows,

$$y = ae^{bx}U \quad (3)$$

With the parameters a and b and the multiplication error term U . If we find the natural logarithm of both sides, this will becomes

$$\ln(y) = \ln(a) + bx + u \quad (4)$$

in which, $u = \ln(U)$

Performance Evaluation

Performance of obtained results with Iron-based modeling approach variation data is evaluated using the root mean square difference (RMSD), mean bias difference (MBD), and Pearson's Coefficient Correlation is expressed by r_{XY} in Equations (5)-(7) respectively, where N is the number of data points, T_i^{Exp} and T_i^{Model} are the i -th ($i = 1, 2, \dots, N$) experimental and model of superconducting transition temperature. $\bar{X}$ and $\bar{Y}$ are mean of a linear relationship between the X_i and Y_i variable of obtained results with Iron-based modeling approach variation data, respectively.

$$RMSD = \sqrt{\frac{\sum_{i=1}^N (T_i^{Model} - T_i^{Exp})^2}{N}} \times 100 \quad (5)$$

$$MBD = \frac{\sum_{i=1}^N (T_i^{Model} - T_i^{Exp})}{\frac{\sum_{i=1}^N T_i^{Exp}}{N}} \times 100 \quad (6)$$

$$r_{XY} = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^N (Y_i - \bar{Y})^2}} \quad (7)$$

Results and Discussion

The actual data (obtained experimentally) employed in this research work was drawn from literatures (Owolabi *et al.*, 2014) as presented in Table 1. The adopted actual data set comprises the true representatives of different families of polycrystalline iron based superconductors (11-type, 111-type, 122-type and 1111-type) prepared using different methods as presented in Table 1. Nonlinear relationships, which are modeled through the SVM from Equation (1) and LRNLT from Equation (4) as presented in Figure 1.

We were used ANOVA together with correlation data to assess the statistical significance of differences in modeling approach variation in Tables 2 and 3, respectively.

In Table 3, correlation coefficient between variations of lattice parameters c ($^{\circ}\text{A}$) and critical temperature T_c (K) were highly significant at the 0.01 level. The meaning of this result describes a perfect direct relationship of a statistical measure of the strength of a linear relationship between variations of lattice parameters c ($^{\circ}\text{A}$) and critical temperature T_c (K). On the other hand, the variations of lattice parameters a ($^{\circ}\text{A}$) were non-significant with critical temperature T_c (K). The meaning of this result describes a perfect direct relationship of a statistical measure of very weak linear relationship between variations of lattice parameters a ($^{\circ}\text{A}$) and critical temperature T_c (K).

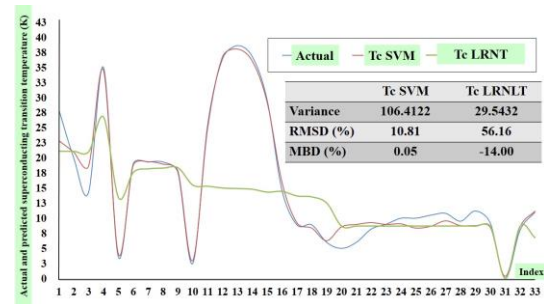


Figure 1. Superconducting transition temperature predictions by SVM and LRNLT

Table 2. Analysis of variance (ANOVA): “Experimental T_c (K)”, “SVM Estimated T_c (K)” and “LRNLT Estimated T_c (K)”

Anova: Single Factor						
SUMMARY						
Groups	Count	Sum	Average	Variance		
Experimental T _c (K)	33	502.20	15.2181	113.2934		
SVM Estimated T _c (K)	33	502.60	15.2303	106.4122		
LRNLT Estimated T _c (K)	33	431.91	13.0881	29.5432		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	100.383	2	50.1915	0.6041	0.5486	3.0911
Within Groups	7975.962	96	83.0829			
Total	8076.345	98	82.4116			

Table 3. The correlation data of experimental T_c (K) through variations of lattice parameters “ a ($^{\circ}\text{A}$)” and “ c ($^{\circ}\text{A}$)”

		a ($^{\circ}\text{A}$)	c ($^{\circ}\text{A}$)	Experimental T_c (K)
a ($^{\circ}\text{A}$)	Pearson Correlation	1	0.430*	0.283
	Sig. (2-tailed)		0.013	0.111
	N	33	33	33
c ($^{\circ}\text{A}$)	Pearson Correlation	0.430*	1	0.520**
	Sig. (2-tailed)	0.013		0.002
	N	33	33	33
Experimental T_c (K)	Pearson Correlation	0.283	0.520**	1
	Sig. (2-tailed)	0.111	0.002	
	N	33	33	33

* Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

However, correlation coefficient between variations of lattice parameters a ($^{\circ}$) and c ($^{\circ}$) were significant at the 0.05 level. Therefore, purpose of the result from Table 3 is a statistical measure of the strength of a linear relationship between two variables (critical temperature T_C (K) through variations of lattice parameter c ($^{\circ}$) and a ($^{\circ}$)) (Taylor, 1990).

In addition, results of significant level from correlation coefficient in Table 3 show that critical temperature T_C (K) through variations of lattice parameter c ($^{\circ}$) and a ($^{\circ}$) possibly due to the layered oxypnictides and the change in the anion height from the Fe layer (Chen *et al.*, 2008; Kamihara *et al.*, 2008). So, we can formulate an optimization correction factor (C.F.) of LRNLT model as follows

$$\ln(y) = \ln(a) + bx + u = A \quad (8)$$

For change in A , rewrite the optimization correction factor from LRNLT model as

$$y = e^A \rightarrow y = e^{C.F.(A)} \quad (9)$$

in which, $C.F.$ is the correction factor.

The creation of an optimization correction factor by using 7th degree polynomial regression forms as follows,

$$\begin{aligned} C.F. = & 3506.2320x^7 - 40128.0276x^6 \\ & + 194388.6399x^5 - 516767.4704x^4 \\ & + 814285.5737x^3 - 760504.1210x^2 \\ & + 389747.768x - 84527.3081 \end{aligned} \quad (10)$$

in which, $x = \ln \frac{c}{a}$

LRNLT from the creation of optimization correction factor, which is created by analyzing a statistical measure of the strength of variation (or variance) relationship between critical temperature

T_C (K) through variations of lattice parameter c ($^{\circ}$) and a ($^{\circ}$). The correction factor was created by the ratio between a highly significant variable (c ($^{\circ}$)) and a very weak significant variable (a ($^{\circ}$)). This methodology corresponds with the idea that variance can be partitioned: within-group variance, was interpreted as computational error and between-group variance, as computational technique (model) variation. Such an idea is required to identify the coefficient of variation (also called R^2) and measure the fraction of the total variation resulting from differences among treatment means (between computational techniques or models) (Krzywinski and Altman, 2014; Armstrong *et al.*, 2022). The result of formulating an optimization correction factor (C.F.) of the LRNLT model was presented in Figure 2 and ANOVA in Table 4, respectively.

The comparisons of obtained results with Iron-based modeling approach by using t-Test as presented in Table 5.

In Figure 2, when an independent variable of “LRNLT (C.F.) Estimated T_C (K)” is continuously measured. By using a validated instrument as the optimization correction factor (C.F.). The raw data so generated closely approximates the true continuous measure more than “LRNLT Estimated T_C ”, having the quality of ratio leveled measured variable from ratio between a highly significant

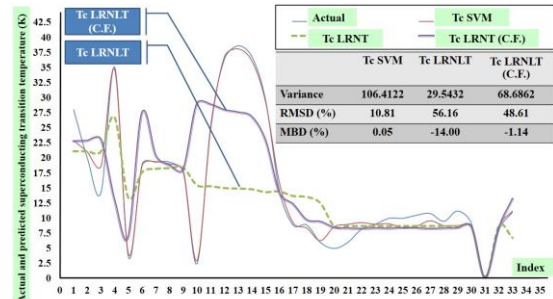


Figure 2. Superconducting transition temperature predictions by LRNLT and LRNLT (C.F.)

Table 4. Analysis of variance (ANOVA): “Experimental T_C (K)”, “SVM Estimated T_C (K)” and “LRNLT (C.F.) Estimated T_C (K)”

Anova : Single Factor SUMMARY						
Groups	Count	Sum	Average	Variance		
Experimental T_C (K)	33	502.20	15.2181	113.2934		
SVM Estimated T_C (K)	33	502.60	15.2303	106.4122		
LRNLT (C.F.) Estimated T_C (K)	33	496.45	15.0439	68.6862		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	0.7176	2	0.3588	0.0037	0.9962	3.0911
Within Groups	9228.5380	96	96.1306			
Total	9229.2550	98	94.1760			

Table 5. t-Test: Two-Sample Assuming Equal Variances

Two-Sample	Pooled Variance	t Stat	df	HMD	t crit
Experimental T_C (K) VS SVM Estimated T_C (K)	109.8528	0.0047	64	0	1.6690
Experimental T_C (K) VS LRNLT Estimated T_C (K)	71.4183	1.0238			
Experimental T_C (K) VS LRNLT (C.F.) Estimated T_C (K)	90.9898	0.0741			
SVM Estimated T_C (K) VS LRNLT Estimated T_C (K)	67.9777	1.0553	64	0	1.6690
SVM Estimated T_C (K) VS LRNLT (C.F.) Estimated T_C (K)	87.5492	0.0809			

Note: t-Test: Two-Sample Assuming Equal Variances at significance level of 0.05, Hypothesized Mean Difference (HMD)

variable (c ($^{\circ}$)) and a very weak significant variable (a ($^{\circ}$)).

This result describes a perfect direct relationship of the basic principles of research design is that of maximizing the systematic or desirable variance, minimizing the error variance and controlling the effect of the extraneous variables. This principle was referred to “MAXMINCON” by Pedhazur and Kerlinger, (1982) and Kerlinger (1986). They agreed that when both dependent and independent variables are continuous, regression analysis should be used in the study of the contribution of the independent variables in the prediction model. When an independent variable is continuously measured as recommended by Nunnally (1978), using a validated instrument, the data so generated closely approximates true continuous measure, having the qualities of ratio leveled measured variable (Kerlinger, 1986).

Underling the definition “MAXMINCON”, within group variance was interpreted as the effect of extraneous variables causing the variations in superconducting transition temperature T_C (K), i.e. valence fluctuation in ions, superexchange mechanism (O’Mahony *et al.*, 2022), oxygen vacancies, oxygen hole content, charge-transfer gap (Kowalski *et al.*, 2021), types of dopants, types of ionic dopant etc. They are not included in between group variance as predictor variable coefficients through different lattice parameters a ($^{\circ}$), c ($^{\circ}$) to estimate the superconducting transition temperature T_C (K). But these factors are included within group variation or “MAXMINCON” of modeling in the ANOVA approach.

The contribution of the block effect to illustrate blocking of variances is shown in Table 2 and Table 4, now a substantial component of the variance in each computational technique. LRNLT (C.F.) estimated T_C (K) has a distribution of variance more than LRNLT estimated T_C (K). But the effect of C.F. can make average T_C (K) the closer Experimental T_C (K) more than LRNLT Estimated T_C (K) corresponds with a decrease in RMSD. This result explains that the development of the LRNLT estimated T_C (K) became LRNLT (C.F.) estimated

T_C (K) with the ANOVA makes the model more accurate near to the famous model, as SVM estimated T_C (K). The reason ANOVA makes the model more accurate is because the result of ANOVA leads to the analysis contribution of the block effect to the deviations (Krzywinski and Altman, 2014).

By using correlation analysis and multiple regression analysis with block effect of the deviations underling the definition “MAXMINCON” in Table 3, it leads to a correlation factor (C.F.) of LRNLT (C.F.) estimated T_C (K) in Equation (10), which causes a decrease in RMSD. To confirm the efficiency of ANOVA, it can reduce the distribution of RMSD and MBD, leading to the model being more accurate for development of the LRNLT model. The decrease in RMSD and MBD was a tendency close to the famous model. ANOVA can be used to test the efficiency between two models from different computational techniques, which can be done using test statistics t as shown in Table 5.

The t stat between Experimental T_C (K) and LRNLT (C.F.) Estimated T_C (K) were lower compared to LRNLT Estimated T_C (K) and be close to SVM estimated T_C (K), at a significant level of 0.05. These results show that LRNLT (C.F.) Estimated T_C (K) started close to non-significant of difference two mean (average T_C (K)) between Experimental T_C (K) and SVM estimated T_C (K) through hypothesized mean difference (HMD) 0 or accept null hypothesis (H_0).

Although the application of ANOVA to develop a simple model (LRNLT) with similar performance to a complex model (SVM) succeeds in reducing the value of RMSD and MBD. But the standard deviation (S.D.) of superconducting transition temperature between experimental T_C (K) and LRNLT (C.F.) estimated T_C (K) was different compared to the standard deviation (S.D.) of superconducting transition between experimental T_C (K) and SVM estimated T_C (K). We assume that such a difference may cause the RMSD and MBD of LRNLT (C.F.) estimated T_C (K) to deviate from experimental T_C (K) rather than SVM estimated T_C (K).

Since, in the view of parametric statistics, ANOVA uses means as representative of the raw data set to analyze standard deviation (S.D.). Such methods may cause the analysis of standard deviation (S.D.) to be influenced by raw data with very large or very small value. Therefore, we propose to combine the median together with ANOVA under the concept of nonparametric statistics. To reduce the influence of extreme raw data and increase the likelihood of RMSD and MBD has decreased, until it is close to SVM for next times.

Conclusions

The application of ANOVA to assess the statistical significance of differences in modeling approach variation is developed to predict doped Fe-based superconductor critical temperature from structural and topological parameters using ANOVA. The high correlation coefficient between the predicted and experimental superconducting transition temperature, the low prediction root mean square difference and mean bias difference, and the outstanding efficiency of the ANOVA results suggest the usefulness of the ANOVA for modeling and understanding the relationship between lattice parameters and superconducting transition temperature.

Acknowledgement

The financial support of Thailand Science Research and Innovation (TSRI) under contract No. 2610293, Pathumwan Institute of Technology is acknowledged.

References

- Armstrong, R.A., Eperjesi, F., and Gilmartin, B. (2002). The application of analysis of variance (ANOVA) to different experimental designs in optometry. *Ophthalmic and Physiological Optics*, 22(3):248-256. <https://doi.org/10.1046/j.1475-1313.2002.00020.x>
- Cai, C.Z., Xiao, T.T., Tang, J.L., and Huang, S.J. (2013). Analysis of process parameters in the laser deposition of YBa₂Cu₃O₇ superconducting films by using SVR. *Physica C: Superconductivity*, 493:100-103. <https://doi.org/10.1016/j.physc.2013.03.038>
- Cai, C.Z., Zhu, X.J., Wen, Y.F., Pei, J.F., Wang, G.L. and Zhuang, W.P. (2010). Predicting the superconducting transition temperature T_c of BiPbSrCaCuOF superconductors by using support vector regression. *Journal of Superconductivity and Novel Magnetism*, 23(5):737-740. <https://doi.org/10.1007/s10948-010-0732-x>
- Chen, X., Dai, P., Feng, D., Xiang, T., and Zhang, F.-C. (2014). Iron-based high transition temperature superconductors, *National Science Review*, 1(3):371-395. <https://doi.org/10.1093/nsr/nwu007>
- Chen, X.H., Wu, T., Wu, G., Liu, R.H., Chen, H., and Fang, D.F. (2008). Superconductivity at 43 K in SmFeAsO_{1-x}F_x. *Nature*, 453(7196):761-762. <https://doi.org/10.1038/nature07045>
- Cortes, C. and Vapnik, V. (1995). Support vector networks. *Machine Learning*, 20:273-297. <https://doi.org/10.1007/BF00994018>
- Diaconis, P. and Efron, B. (1983). Computer-intensive methods in statistics. *Scientific American*, 248:116-130. <https://doi.org/10.1038/scientificamerican0583-116>
- Efron, B. and Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, 37(1):36-48. <https://doi.org/10.1080/00031305.1983.10483087>
- Goldman, L., Caldera, D.L. and Nussbaum, S.R. (1977). Multifactorial index of cardiac risk in noncardiac surgical procedures. *The New England Journal of Medicine*, 297:845-850. <https://doi.org/10.1056/NEJM197710202971601>
- Goldman, L., Debra, M.P.H., Caldera, L., Samuel, R.N., Nussbaum, R., Southwick, F.S., Krogstad, D., Murray, B., Burke, D.S., O'Malley, T.A., Goroll, A.H., Caplan, C.H., Nolan, J., Carabello, B., and Slater, E.E. Gupta, S.M. (2007). Support vector machines based modelling of concrete strength. *World Academy of Science, Engineering and Technology*, 36:305-311.
- Hamidieh, K. (2018). A data-driven statistical model for predicting the critical temperature of a superconductor. *Computational Materials Science*, 154:346-354. <https://doi.org/10.1016/j.commatsci.2018.07.052>
- Kamihara, Y., Watanabe, T., Hirano, M., and Hosono, H. (2008). Iron-based layered superconductor La[O_{1-x}F_x]FeAs ($x = 0.05-0.12$) with $T_c = 26$ K. *Journal of the American Chemical Society*, 130(11):3296-3297. <https://doi.org/10.1021/ja800073m>
- Kerlinger, F.N. (1986) *Foundations of Behavioral Research*. 3rd ed. Holt, Rinehart and Winston, NY.
- Kowalski, N., Dash, S.S., Sémon, P., and Tremblay, A.-M. (2021). Oxygen hole content, charge-transfer gap, covalency, and cuprate superconductivity. *PNAS*, 118(40):e2106476118. <https://doi.org/10.1073/pnas.2106476118>
- Krzywinski, M. and Altman, N. (2014). Analysis of variance and blocking. *Nature Methods*, 11(7):699-700. <https://doi.org/10.1038/nmeth.3005>
- Nunnally, J.C. (1978). *Psychometric Theory*. 2nd ed. McGraw-Hill, NY.
- O'Mahony, S.M., Ren, W., Chen, W., Chong, Y.X., Liu, X., Eisaki, H., Uchida, S., Hamidian, M.H., and Davis, J.C.S. (2022). On the electron pairing mechanism of copper-oxide high temperature superconductivity. *PNAS*, 119(37):e2207449119. <https://doi.org/10.1073/pnas.2207449119>
- Owolabi, T.O., Akande, K.O. and Olatunji, S.O. (2014). Prediction of superconducting transition temperatures for Fe-based superconductors using support vector machine. *Advances in Physics Theories and Applications*, 35:12-26.
- Owolabi, T.O., Akande, K.O. and Olatunji, S.O. (2016). Application of computational intelligence technique for estimating superconducting transition temperature of YBCO superconductors. *Applied Soft Computing*, 43:143-149. <https://doi.org/10.1016/j.asoc.2016.02.005>
- Owolabi, T.O., Akande, K.O., and Olatunji, S.O. (2015). Estimation of superconducting transition temperature T_c for superconductors of the doped MgB₂ system from the crystal lattice parameters using support vector regression. *Journal of Superconductivity and Novel Magnetism*, 28:75-81. <https://doi.org/10.1007/s10948-014-2891-7>
- Pedhazur, E.J. and Kerlinger, F.N. (1982). *Multiple Regression in Behavioral Research: Explanation and Prediction*. 2nd ed. Holt Rinehart and Winston, NY.

-
- Taylor, R. (1990). Interpretation of the correlation coefficient: a basic review. *Journal of Diagnostic Medical Sonography*. 6(1):35-39. <https://doi.org/10.1177/875647939000600106>
- Vapnik, N.V. (1995). *The Nature of Statistical Learning Theory*, Springer, 188p. <https://doi.org/10.1007/978-1-4757-2440-0>
- Zhang, Y. and Xu, X. (2021), Predicting doped Fe-based superconductor critical temperature using machine learning. *International Journal of Materials Research*, 112(1):2-9. <https://doi.org/10.1515/ijmr-2020-7986>